DOCUMENT RESUME

ED 137 400 TM 006 206

AUTHOR Gorham, William A.
TITLE Some Psychometric and Legal Considerations in Cooperative Testing. Professional Series 74-4.
INSTITUTION Civil Service Commission, Washington, D.C. Personnel Measurement Research and Development Center.
PUB DATE Nov 74
NOTE 11p.; Paper presented at the meeting of the International Personnel Management Association (October 15, 1974)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS City Government; Federal Government; Federal Legislation; *Government (Administrative Body); *Interagency Cooperation; Job Analysis; *Legal Problems; Local Government; *Occupational Tests; Policy Formation; *Psychometrics; State Government; Test Construction; Test Reliability; Test Results; Test Validity
IDENTIFIERS Intergovernmental Personnel Act

ABSTRACT

Cooperative testing is a rather specific and practical aspect of validity generalization or test transportability. It refers to two or more units of government combining their testing and/or recruiting efforts. The Intergovernmental Personnel Act of 1970 (P.L. 1-648) made specific provision for cooperative examining in stating that the Civil Service Commission may join, on a shared-cost basis, with State and local governments in cooperative recruiting and examining activities under agreed upon procedures and regulations. Cooperative testing can also be thought of as the referral by one government to another of any names prescreened by any test or qualification. The need for broadening the concept of cooperative testing became evident as review of existing and proposed cooperative agreements raised psychometric, legal, and policy questions concerning the interchange of test scores, including the extent of Federal responsibility for ensuring the job relatedness and fairness of tests used for more than one government. This review took place in the spring of 1973 when a task force of operating, legal, and psychometric experts met within the Civil Service Commission. The results of that review, as they relate to cooperative testing, are the basis for this paper. (Author/MV)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

PRDC Personnel Research and Development Center

Professional Series 74–4

ED137400

Some Psychometric and Legal Considerations in Cooperative Testing

U.S. DEPARTMENT OF HEALTH EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

United States
Civil Service Commission
Bureau of Policies and Standards

2

IM 008 200

Professional Series PS-74-4

SOME PSYCHOMETRIC AND LEGAL CONSIDERATIONS IN COOPERATIVE TESTING

(Paper delivered as part of a symposium, "Portability of Test Validation." International Personnel Management Association, October 15, 1974.)

William A. Gorham

U.S. Civil Service Commission
Personnel Research and Development Center
Washington, D.C.
November 1974

Cooperative testing is a rather specific and practical aspect of validity generalization or test transportability. It refers to two or more units of government combining their testing and/or recruiting efforts. Often the reasons for cooperative testing are worthy: to conserve testing and technical resources; public convenience, that is making it easier for applicants to get a job by taking a single test instead of two or more; and to speed up hiring and placement. As we are aware, it is not uncommon for up to four governmental jurisidctions to be recruiting and examining in the same area— Federal, State, county, and city. Sometimes special jurisdictions are also involved. All may be competing for a similar labor market.

The Intergovernmental Personnel Act of 1970 (P.L. 1-648) made specific provision for cooperative examining:

> "The Commission (U.S. Civil Service Commission) may join, on a shared-cost basis, with State and local governments in cooperative recruiting and examining activities under such procedures and regulations as may jointly be agreed upon."

Cooperative testing can also be thought of as the referral by one government to another of any names prescreened by any test or qualification. The need for broadening the concept of cooperative testing became evident as review of existing and proposed cooperative agreements raised psychometric, legal, and policy questions concerning the interchange of test scores, including the extent of Federal responsibility for ensuring the job relatedness and fairness of tests used for more than one government. This review took place in the spring of 1973 when a task force of operating, legal, and psychometric experts met within the Civil Service Commission. The results of that review, as they relate to cooperative testing, have served as the basis for this paper.

While I have cast the discussion around the Federal government as one of the cooperators in any potential agreement, obviously it does not have to be so. The general psychometric and legal considerations will, I believe, still obtain if a cooperative arrangement is worked out between or among non-Federal governments.

While there are a number of assumptions which should underpin cooperative testing arrangements, a key one is that the government whose test is being used has a responsibility for the manner in which that test and the results of the test are used by other jurisdictions. It must maintain reasonable assurance that procedures

and practices in the use of tests and their results are consonant with the intent of its own policies and standards.

If, for example, Federal tests are involved, there would appear to be varying levels or degrees of Federal responsibility (and risk) associated with various possible cooperative situations. Listed in order of decreasing Federal involvement and responsibility for ensuring how tests and scores are used, the situations might be ranked very generally as follows:

1. Federal government uses Federal test and/or State or local test

2. Federal government uses State test

3. State or local government uses its own test and/or Federal test (i.e. State or local accepts Federal scores, but Federal does not accept State or local scores).

4. State or local government uses Federal test

5. Federal and State or local government use a third test.

Each of the above situations in turn has variations with associated degrees of risk, such as who administers the tests, and whether separate or joint registers are maintained; but at this point it would appear that *these* factors are to some degree involved in all cooperative testing situations and could thus be treated at a general level.

At least four areas of consideration are of concern in those types of situations in which exchange of test scores is involved. These are:

(1) Job Analysis

(2) Test Plan Similarity

(3) Visual Similarity

(4) Psychometric Similarity

A discussion of these considerations will, I believe lead to conclusions about the viability of various cooperative arrangements.

JOB ANALYSIS

If a cooperative arrangement is to be undertaken, each user must have a job analysis of the position or job in question. At a minimum the job analysis must describe the job and the necessary knowledges, skills, abilities and other worker characteristics necessary or important for effective job performance.

A problem is that the procedures for conducting job analyses and the bases for establishing the relationship between the test and the job are not standardized and are open to various avenues of approach. The lack of standardized procedures for conducting job analyses and documenting the job relatedness of a test for employment does not in itself stand as a barrier to cooperative testing arrangements, but it does make cooperative testing more difficult.

If a universally accepted set of job-analysis equivalency procedures existed in one or more documented sources, criteria could be applied fairly readily as to the comparability of jobs in different jurisdictions. Some such procedures exist. Obviously those which are preferred have quantitative solutions to measuring job similarity. Optimally, cooperating jurisdictions should use the same method. In the absence of quantitative procedures, judgments must be applied to job analyses. It would seem that at a minimum, experts from the cooperating jurisdictions should concur that the jobs are alike, and that the basis for these judgments should be documented. It would of course be desirable that the judgments were made independently, but using criteria agreed upon in advance. This would ensure that there is at least professional concurrence between cooperating jurisdictions that the jobs are enough alike to proceed with looking at other aspects of a cooperative agreement.

TEST PLAN COMPARABILITY

If more than one test is to be used in the cooperative agreement, there must be test plan comparability.

A test plan is essentially the specifications for test construction. As indicated by Thorndike (1949, p.50), the specifications cover such points as the following:

(1) The function or functions that the test is to measure

(2) Illustrations of each type of item to be included in the test

(3) The number of each type of item to be included.

(4) The range of content to be covered in the test, where variety of content is a factor, and the allocation of items within that range.

(5) The time limits for the test, or for each separately timed section of it.

(6) The nature of the population for which the test is designed.

(7) The desired level and range of difficulty of test items for the population.

(8) The editorial and statistical procedures to be used in selecting and refining test items.

These specifications help ensure that the test and any alternate forms meet intended requirements in terms of functions measured, scope and difficulty.

If we assume, for purposes of illustration, that test plans are in hand for two tests, and that these tests have been developed independently for the same function(s), it is extremely unlikely that the tests would be alike on other specifications. These differences in specifications will lead to a difference in performance by the same person taking both tests, aside from reliability differences between the tests. There is usually a difference (other than reliability) in scores when an individual takes two tests when the tests are based on the *same* test plan; but these differences will be much smaller, can be adjusted for on the average, and would not reflect a systematic uncontrolled bias.

Differences in test plans have implications for exchangeability of scores. Since differences in test plans will almost inevitably lead to differences in psychometric characteristics in test application, it is perhaps worthwhile to consider closely the advisability of expending resources to conduct comparability analyses based on special test administrations, when it is known that the test plans differ.

No procedures exist for satisfactorily measuring differences in test plans, for projecting the specific impact of differences, or for determining the amount of difference

which can be reasonably tolerated, when the intent is to treat test scores as equivalent.

VISUAL SIMILARITY

Visual similarity of tests is closely related to *face validity* and to test plan similarity. Face validity refers to the apparent applicability of a test for the use to which it is put. The apparent applicability or relevance of the test can vary depending upon the observer, e.g., applicant, psychologist, personnel manager, judge or jury. Thus, face validity may be associated in varying degrees with the actual validity, or value of the test for its intended use. More importantly, face validity can strongly influence the decisions of those who may challenge the test on some grounds, as well as those who have the power to decide whether or not a test can be used. In cases where there are reservations concerning psychometric factors or job relatedness, the practical importance of face validity increases, but does in no way substitute for technical weaknesses of a test.

When tests intended for the same purpose are compared, the possibility of challenge to either or both of the tests increases. As the above discussion on test plan similarity implies, visual comparison of tests will highlight differences which may exist in apparent content, types of items, type and number of response alternatives, time allowances, difficulty level, and other characteristics.

These considerations indicate caution in the use of tests which may operate the same, when they appear different. A good case may be made that when tests look different, they *are* different--at least for purposes of exchangeability.

NECESSITY FOR VALIDITY

The validity of a test must be demonstrated in at least one professionally acceptable way. The test user has the responsibility for demonstrating validity for the jobs for which he uses the test.

Given that validity of a test is documented for certain jobs in one government jurisdiction, another government may use the test for similar jobs and meet its validity responsibility by relying on the validity study previously done for

a similar job. Similar jobs are jobs which have similar duties or require similar skills, knowledges, or abilities.

New validity evidence is necessary, however, if the test is to be used for jobs which are determined on the basis of job analysis to be different from those for which validation has been conducted, or if the situations are markedly different. This principle is implicit in the basic meaning of validity, i.e., a test's value for the use to which it is put. Thus, the requirement for new validity studies in conjunction with any proposed cooperative agreement will depend upon (a) whether adequate validity documentation exists for the current use of the test, and (b) whether the extended application of a test to another jurisdiction involves similar or different jobs.

SITUATIONS INVOLVING COOPERATIVE USE OF MULTIPLE TESTS

Where the cooperative arrangement involves any government using its own test and that of another government interchangeably, the requirements to ensure equitable merit examining are greater than in a single test situation. As a first requirement it is incumbent upon each jurisdiction to meet for both tests the job analysis and validity requirements as set out above. Secondly, the tests involved must be so similar as to be approximately equivalent, or parallel, forms of the same test. This requirement is based upon a number of considerations, which follow below.

In the multiple test situation scores of applicants taking different tests must be merged to form a ranked list. Conversion formulas can be developed for converting scores from one test to "comparable" scores on another test. Such conversions entail a significant amount of error--unless the tests are essentially parallel or equivalent forms. Unless two tests are built to be equivalent, i.e., based on the same test plan (test specifications), they will vary in *what* they measure and *how well* they measure it. These differences are not of practical concern as long as the tests are used separately . Differences between a State's clerical test, for example, and a Federal clerical test are not a cause for concern as long as each is applied only within its jurisdiction. Both may meet requirements for job relatedness and validity for jobs in either or both jurisdictions. When test scores from these tests are merged, however, the differences become of significant concern. The minor

differences in *what* the tests measure and *how well* they measure it may become compounded with the resulting effect that an individual may rank quite differently in the resulting list depending upon which test he takes and whether he scores high or low. Since the errors are over and above those normally entailed in any single test procedure, psychometric authorities caution against the use of conversion (transmutation) tables between tests for selection purposes (although such tables are highly useful for counseling purposes).

In summary if the tests involved are not very close to alternate forms, or if they lack adequate job analysis or validity support for use in any of the government jurisdictions involved, the use of conversion tables would be improper. The risks of such an agreement are high since a legal decision against one jurisdiction would likely affect the other. Because of the complex nature of such agreements, it is doubtful if many viable cooperative arrangements are possible, likely, or desirable.

SITUATIONS INVOLVING COOPERATIVE USE OF ONE TEST

Where the cooperative arrangement involves only one test, (State uses Federal test, Federal uses State test, or both use a third test) the situations are quite similar in what requirements are necessary, and it is in these situations that cooperative testing is most viable. Similarity of jobs must still be demonstrated, but that hurdle passed, the complex psychometric considerations are reduced to those required of a single test: Objectivity, reliability and validity.

COOPERATIVE REGISTERS

Finally, let us turn to another aspect of the IPA, in which the Commission also may, on the written request of a State or local government and under such procedures as may be jointly agreed upon, certify to such governments from appropriate Federal registers the names of potential employees.

I would be less than candid if I did not tell you that elements of cooperative testing are present in this situation, and that opinion is divided on the responsibilities of the cooperating jurisdictions.

If such referrals constitute, in effect, certificates whereby these registers become the sole source of examination and appointment, their use should be contingent upon the user's showing of job relatedness and validity for the positions for which they are used, for the examining instruments involved.

This approach is a substantial change in current policy with regard to the use of Federal registers. Present procedures make no distinction as to whether State and local government use is as a general and supplemental recruiting source or as a source of certification. The Commission has not imposed, in either case, a job relatedness and validity test as a condition of use.

The above change would clearly distinguish, procedurally, between the use of these registers for recruitment purposes and their use under conditions tantamount to certification. One view holds that recruitment use would not, and should not, entail Commission involvement in job relatedness and validity determinations. This responsibility would clearly remain with the State or local government. But, register use which would make the Commission responsible for rating and ranking of candidates would require that it become involved and assume responsibility for the job relatedness and validity of the pertinent examining and certification procedures.

Another view holds that the Commission must assume the same responsibility in both situations. Since names are put on a CSC register after a weeding out process, making such names available to State and local governments as one of many potential sources of recruitment is, as a matter of law, the same as certification from a CSC register, even if the State or local government applies its own applicant appraisal, ranking, certification and selection procedures.

My personal view is that there is a great deal of self-selection in applying for jobs and that to the degree that this results in a narrowed ability range, the onus *may* fall on the government which allows its registers to be used to demonstrate the job-relatedness of the procedures for the borrowing government's jobs. If it is not willing to do that, it should not be involved in cooperative register referrals. Where applications are merely received, however, with little narrowing or screening, there will, from a practical point of view, be little responsibility except to the user to document job relatedness.

GPO 885-869